



The effect of a coding strategy that removes temporally masked pulses on speech perception by cochlear implant users

Wiebke Lamping ^{a, b, *}, Tobias Goehring ^b, Jeremy Marozeau ^a, Robert P. Carlyon ^b

^a Hearing Systems Section, Department of Health Technology, Technical University of Denmark, DK-2800, Kgs. Lyngby, Denmark

^b Medical Research Council Cognition and Brain Sciences Unit, University of Cambridge, 15 Chaucer Road, Cambridge, CB2 7EF, United Kingdom

ARTICLE INFO

Article history:

Received 16 September 2019

Received in revised form

26 March 2020

Accepted 5 April 2020

Available online 10 April 2020

Keywords:

Cochlear implant

Speech perception

Temporal integrator

Masking

ABSTRACT

Speech recognition in noisy environments remains a challenge for cochlear implant (CI) recipients. Unwanted charge interactions between current pulses, both within and between electrode channels, are likely to impair performance. Here we investigate the effect of reducing the number of current pulses on speech perception. This was achieved by implementing a psychoacoustic temporal-masking model where current pulses in each channel were passed through a temporal integrator to identify and remove pulses that were less likely to be perceived by the recipient. The decision criterion of the temporal integrator was varied to control the percentage of pulses removed in each condition. In experiment 1, speech in quiet was processed with a standard Continuous Interleaved Sampling (CIS) strategy and with 25, 50 and 75% of pulses removed. In experiment 2, performance was measured for speech in noise with the CIS reference and with 50 and 75% of pulses removed. Speech intelligibility in quiet revealed no significant difference between reference and test conditions. For speech in noise, results showed a significant improvement of 2.4 dB when removing 50% of pulses and performance was not significantly different between the reference and when 75% of pulses were removed. Further, by reducing the overall amount of current pulses by 25, 50, and 75% but accounting for the increase in charge necessary to compensate for the decrease in loudness, estimated average power savings of 21.15, 40.95, and 63.45%, respectively, could be possible for this set of listeners. In conclusion, removing temporally masked pulses may improve speech perception in noise and result in substantial power savings.

© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Speech recognition in background noise, and sometimes even in quiet, is still difficult for cochlear implant (CI) recipients (Zeng et al., 2008). In addition to the limited number of electrodes, the shallow insertion depth of the array and differences in local neural survival, one major cause for the poor spectral and temporal resolution of CI users is the spread of current along the cochlea. This current spread causes broad regions of the auditory nerve to be stimulated when activating a single electrode, and numerous psychophysical studies have demonstrated substantial interactions when activating two or more channels (Marozeau et al., 2015; McKay et al., 2001; McKay and McDermott, 1996; Shannon, 1983; Townshend et al., 1987).

The overall stimulation pattern delivered to the recipient

depends on the incoming sound, the non-linear mapping from acoustic to electric stimulus (Shannon et al., 2004) and the coding strategy (Wouters et al., 2015). Contemporary CI processing strategies, such as the HiRes120 (Advanced Bionics Corporation) or the FSP (Med-El Corporation), are based on Continuous Interleaved Sampling (CIS, Wilson et al., 1991) using all available channels for stimulation in each time frame. However, for CI listeners speech perception does not seem to improve markedly beyond about 4–8 channels (Dorman et al., 1997; Friesen et al., 2001; Fu and Nogaki, 2005), which is below the number of electrodes of the implant, ranging from 12 to 22 depending on the manufacturer. Thus, reducing the number of active electrodes within a short time frame may in principle decrease channel interactions and improve spectral resolution. However, with the Advanced Combination Encoder (ACE) strategy (Cochlear Corporation) where n out of m (n -of- m) channels, typically 8 out of 22, with the highest amplitudes are selected for stimulation in each time frame (McDermott et al., 1992; Vandali et al., 2000), only small improvements in speech perception could be achieved over CIS (Skinner et al., 2002). As those

* Corresponding author. Department of Clinical Neurosciences, University of Cambridge, Cambridge, CB2 0SZ, United Kingdom.

E-mail address: wl383@cam.ac.uk (W. Lamping).

maxima are likely to be assigned to adjacent channels, more recent strategies have been proposed that potentially decrease interactions by avoiding the activation of neighboring channels in n-of-m strategies (e.g. based on spectral masking), thereby increasing the spectral contrast (Buechner et al., 2008; Nogueira et al., 2016, 2005). Improvements in speech intelligibility could be achieved over a standard n-of-m strategy but remained small, and listeners still struggled to understand speech even in modest levels of background noise. A commercial implementation of a spectral-masking based strategy revealed no benefit in speech perception but decreased power consumption compared to ACE (Buechner et al., 2011).

Even though channel interactions may be reduced by strategies like ACE or by increasing the distance between active channels used for stimulation, current spread with monopolar stimulation is still sufficiently broad to result in a smeared representation of the stimulus arriving at the spiral ganglion neurons (SGNs). Thus, for multi-channel stimulation, the SGNs will receive current not only from electrodes in their vicinity but also from those that are remote, causing the firing pattern of the neurons to be affected by both spectral and temporal interactions (Boulet et al., 2016). These temporal interactions can occur not only at supra-threshold levels (e.g. refractoriness, adaptation) but even when current is received at a subthreshold level. For instance, nerve excitability has been shown to be increased by subthreshold stimulation when inter-pulse intervals are very short, commonly referred to as temporal summation or facilitation (Bierer and Middlebrooks, 2004; Heffer et al., 2010). Facilitation can also be observed in CI listeners using single-electrode stimuli as the detection of a single pulse can be enhanced by a preceding pulse (Cosentino et al., 2015; Nelson and Donaldson, 2001). The influence of subthreshold pulses on speech perception remains unclear, but unwanted charge interactions in the temporal domain may lead to further distortions of the stimulus envelope.

When decreasing the signal carrier rate on which the speech envelope is imposed, and thereby decreasing the number of overall current pulses that could interact with one another, some studies have reported an improvement in intelligibility or higher preference when using low carrier rates to deliver the speech signal (Balkany et al., 2007; Brochier et al., 2017; Park et al., 2012; Vandali et al., 2000). However, other studies demonstrated no benefits of low rates (Plant et al. 2002, 2007; Skinner et al., 2002; Friesen et al., 2005; Weber et al., 2007; Arora et al., 2009; Shannon et al., 2011), while again others have shown that speech intelligibility is better when high carrier rates in the range above 1000 pulses per second (pps) are used (Kiefer et al., 2000; Loizou et al., 2000; Nie et al., 2006; Verschuur, 2005). Further, for most of the above-mentioned studies, a large across-subject variability was found. The lack of effect of carrier rate across studies may be due to a trade-off: high rates can provide better sampling of the envelope (Wilson et al., 1988) and have also been shown to result in more stochastic firing of neurons (Rubinstein et al., 1999), which may be beneficial because neural responses are considerably stronger than normal with electric stimulation (Dynes and Delgutte, 1992). However, these benefits may be offset by the increase in temporal and spectral interactions (McKay et al., 2005; Middlebrooks, 2004).

Reducing spectral and temporal interactions could lead to improved speech perception in noise, but the interplay between current pulses is complex (e.g. Langner et al., 2017). The wide range of outcomes when varying carrier rate or channel selection strategy suggests that a more refined approach to deliver the stimulus pattern is necessary. The processing strategy that is introduced in the following, the temporal integrator processing strategy (TIPS), aims to reduce interactions between current pulses by removing those pulses that are masked in the temporal domain. A

phenomenological model based on psychophysical masking, the sliding temporal integrator (TI), was implemented to estimate which current pulses are less likely to be perceived by the listener. The TI model can account for various temporal effects in normal-hearing listeners, from temporal resolution and masking, to loudness changes and modulation detection (for an overview, see Moore, 2007). Variations of the TI model have previously been used to model aspects of temporal processing by CI users, such as the effect of inter-pulse intervals and amplitude modulation on loudness (McKay and Henshall, 2010; McKay and McDermott, 1998), and the effect of pulse rate and pulse duration on masked and unmasked detection thresholds (Carlyon et al., 2005; McKay et al., 2013; Shannon, 1989).

When applied to acoustic hearing, the TI model incorporates a bank of bandpass filters, followed by a compressive stage that takes the nonlinearity of the cochlear periphery into account. Thus, the integration window could be considered a linear smoothing process acting upon the intensity of the vibration of the basilar membrane and being linearly related to the auditory nerve firing rate (Moore et al., 1996; Oxenham, 2001; Oxenham and Moore, 1994; Plack et al., 2002; Plack and Moore, 1990; Plack and Oxenham, 1998). Subsequently, McKay et al. (2013) used the TI window directly on estimated nerve responses, summing the neural activity within the window to model aspects of perception by CI listeners. Using this approach, they could successfully predict the effects of inter-pulse intervals on detection thresholds and loudness growth, temporal modulation transfer functions, the effect of duration on detection thresholds, and forward masking decay for CI listeners and users of an auditory midbrain implant.

Here the TIPS strategy only includes two stages of the TI model, which are the TI window, acting upon the current pulses, and a decision device, which used a criterion on the TI window output to detect masked pulses. The experiments described in the following investigate the effect of removing current pulses, based on the TI model, on speech intelligibility in quiet and in noise for CI listeners. Further, as fewer current pulses could be used for stimulus delivery, this strategy may also reduce power consumption. Thus, potential power savings were estimated for the test conditions with removed pulses, while taking into account the increase in charge that was required to achieve equal loudness across conditions.

2. Methods

2.1. Overview of the TI processing strategy

The TIPS strategy removes temporally-masked current pulses by implementing a sliding temporal integrator (e.g. Oxenham and Moore, 1994; Plack et al., 2002) and embedding it into an experimental CIS processing scheme. For TIPS, the input frequency range for filtering, envelope extraction and nonlinear acoustic-to-electric mapping were similar to contemporary CI processing schemes and, as noted above, only the TI window and the decision device were added as new elements to the processing chain, which is shown in Fig. 1.

The TI window is generally modeled as a pair of back-to-back exponential functions, to account for both forward and backward masking:

$$W(t) = \begin{cases} (1-r)e^{\left(\frac{t}{b_1}\right)} + r e^{\left(\frac{t}{b_2}\right)}, & t < 0 \\ e^{\left(\frac{-t}{a}\right)}, & t \geq 0 \end{cases}$$

where $W(t)$ is the window shape, or weighting function, at time t

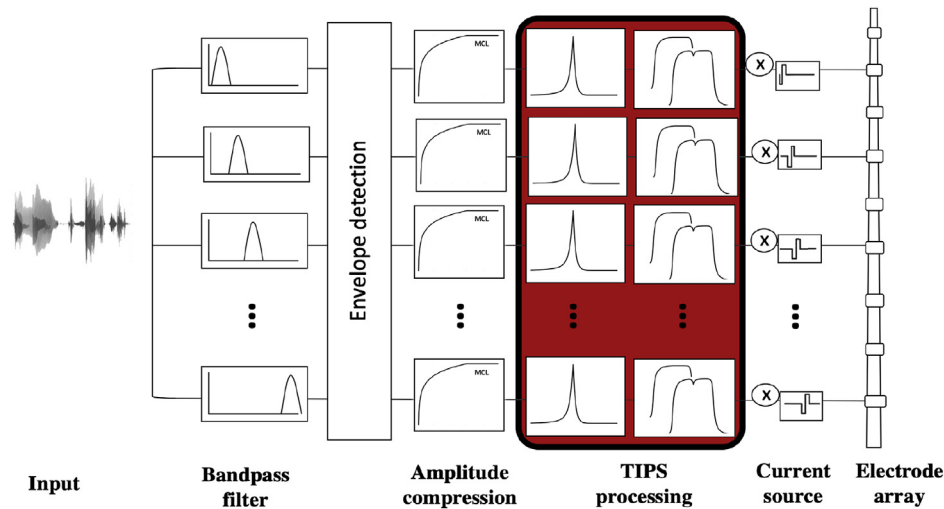


Fig. 1. Block diagram and signal processing of TIPS; after regular bandpass filtering, envelope extraction and amplitude compression, the TI window and decision device were added to the processing chain.

relative to the center of the window function. The parameter r is the weighting of the time constants T_{b1} and T_{b2} for the negative side of the exponential associated with forward masking, while T_a is the time constant for the positive part of the exponential, which is associated with backward masking (Oxenham, 2001; Plack et al., 2002). The integration window shape that provided the best fit to forward masking data in normal-hearing listeners was reported with parameters set to $T_a = 3.5$ ms, $T_{b1} = 4.6$ ms, $T_{b2} = 16.6$ ms, and $r = 0.17$ (Oxenham, 2001) which were thus used by McKay et al. (2013) and also for TIPS.

The input to the TI model consisted of electric stimuli, i.e. current pulses, which were obtained by processing an acoustic signal with a CI processing strategy emulation from the Nucleus Matlab Toolbox (NMT, Swanson and Mauch, 2006). The output of the TI window was the weighted average of the signal over a time interval with an equivalent rectangular duration of about 7 ms, which was then fed to a decision device. A decision criterion was applied to the log-transformed maximum difference between the output of the TI window for the original stimulus and the output of the TI window for that same stimulus minus one pulse at the center of the window, where the highest weight of the window function occurs (see Fig. 2, left). The decision criterion to predict forward masking is commonly set to 3 dB, based on studies with normal-hearing listeners (Plack and Oxenham, 2002). However, temporal effects, such as forward-masked thresholds, are considerably subject- and electrode dependent in CI listeners. The decision criterion was therefore used as an experimental parameter; to test the effect of removing pulses over a wide range, the criteria were set to 1, 1.3, and 1.8 dB to remove 25, 50, and 75% of current pulses, respectively. These values were determined as the average percentage of pulses removed for 10 sentences from the speech corpus that was used for the speech intelligibility tests (see section 2.2.4). Fig. 2 (right) shows the percentages of pulses removed with a range of values of the decision criterion up to 3 dB for an example sentence.

Keeping the criterion fixed, each channel was processed independently by the sliding temporal integrator and the decision device. If a pulse was removed during this process, the stimulus was updated for the ongoing analysis and, after all channels were processed, the signal was transformed back into a stimulus pattern that could be streamed to the implant.

Fig. 3 shows comparisons for the electric stimulation strategy of a sentence for a CIS (0%) reference (Fig. 3a) and three TIPS

conditions where 25, 50, and 75% of current pulses were removed from the reference (Fig. 3b–d). The parameter settings for the stimulation pattern that can be seen in these graphs will be explained in section 2.2.2. Fig. 3 shows that despite removing a substantial number of current pulses, the stimulus envelope seems well maintained. Fig. 4 shows comparisons between 0% (left) and 50% (right) of pulses removed for a short segment of the same sentence as used for Fig. 3, during the word “the”, on electrode 20 (top) and for the vowel /a/ (first phoneme from syllable “asa”, bottom), again on electrode 20. It is worth noting that pulses are predominantly removed just before the onset and just after the offset of the stimulus (see top panel), as well as in the troughs of the envelope modulations (see bottom panel). It is possible that removing pulses in the troughs could reduce the amount of envelope information available to the listener. However, recent evidence suggests that CI listeners are largely insensitive to the pattern of modulation in the envelope troughs (Monaghan et al., 2019). Overall, TIPS seems successful in removing substantial amounts of temporally-masked current pulses based on the TI model.

2.2. Experiments

2.2.1. Subjects

Eight post-lingually deafened native Danish speakers took part. Their mean age was 66.3 years, three were male, and all were users of a Cochlear™ device. Subjects had at least one year of experience with their device and the ACE processing strategy. Their biographical data are given in Table 1. Data were collected at the Technical University of Denmark, and the research was approved by the Science-Ethics Committee for the Capital Region of Denmark (reference H-16036391). All subjects provided written informed consent before taking part. For each individual, impedances were monitored at the beginning and at the end of each session, and it was ensured that the voltage requirements were always kept below the compliance limit for electrical stimulation. Speech intelligibility scores with clinical settings were not available for this set of participants.

2.2.2. Stimuli and map settings

The CIS strategy emulation of the NMT was used to transform the acoustic signals to electric stimuli, which were then streamed to the recipient's implant using the Nucleus Implant Communicator

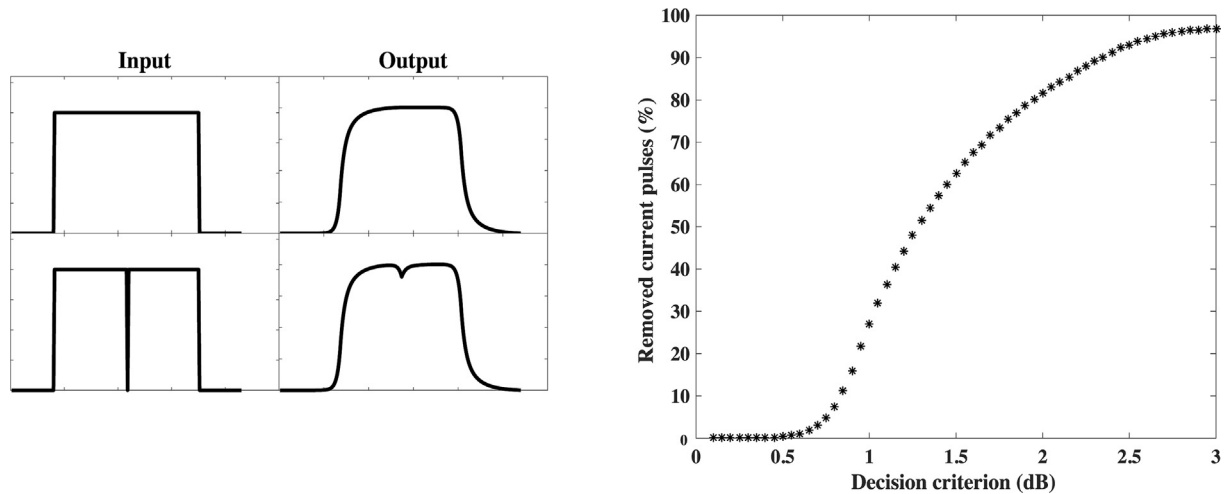


Fig. 2. Left: Input to and output of the temporal integrator for a schematic stimulus (top) and the same stimulus minus one pulse (bottom). Right: Percentage of pulses removed across different criteria for the decision device for the example sentence “Ingrid finder syv røde huse”.

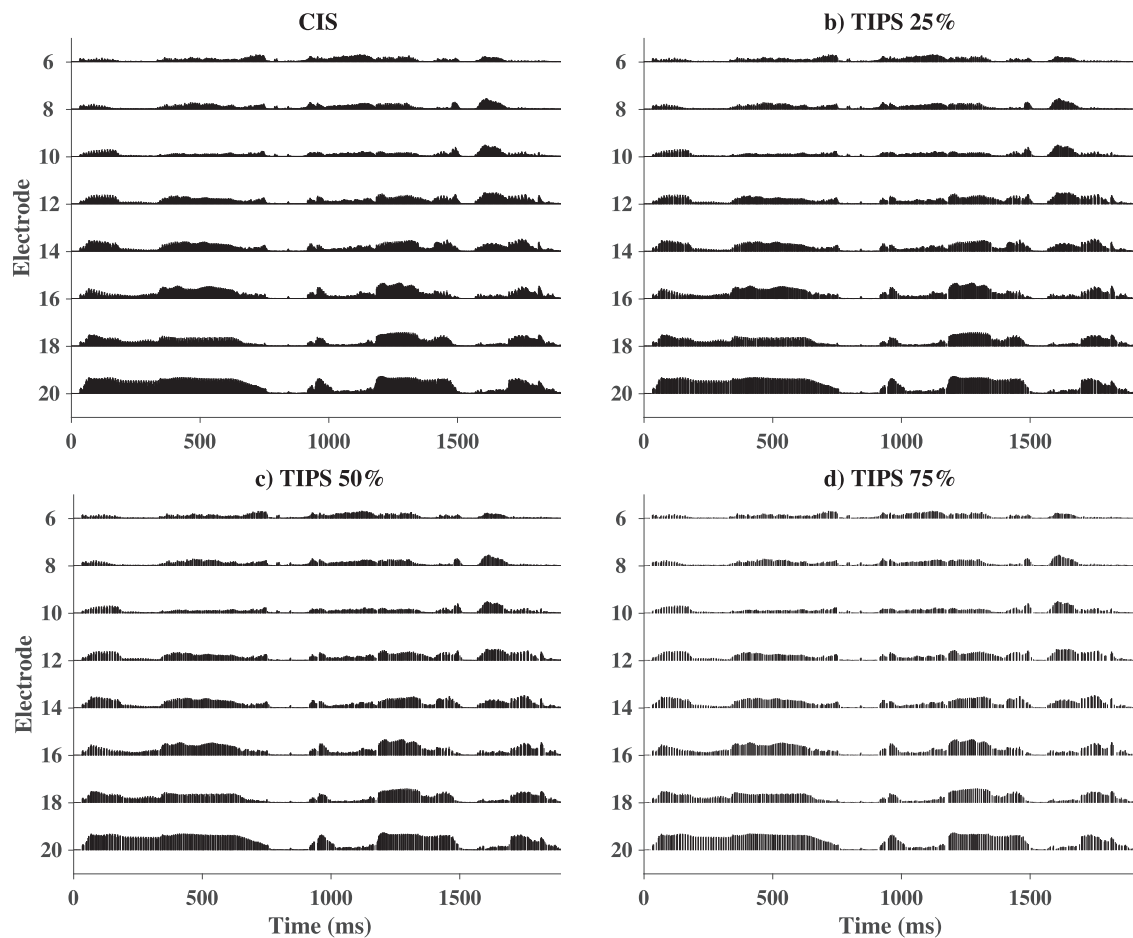


Fig. 3. Example of stimulus pattern for the sentence “They moved the furniture” with a) CIS reference map using only 8 electrodes. The same reference map was then used to generate b) TIPS 25%, c) TIPS 50% and d) TIPS 75%. The dynamic range was set to be 245 clinical units.

(NIC) and the CP920 research processor, both provided by the Cochlear Corporation (Sydney, Australia). The tests were performed using custom Matlab (The MathWorks, Inc., Natick, MA, US) experimental interfaces, modified to incorporate the NIC.

For the experiments, a CIS reference map was generated with a

pre-defined number of fixed electrodes dispersed across the electrode array (see Fig. 3a). CIS was used instead of ACE, which was the regular processing strategy for all subjects, so as to provide an approximately equal amount of acclimatization across conditions. Electrodes, 20, 18, 16, 14, 12, 10, 8, and 6 were selected for stimulus

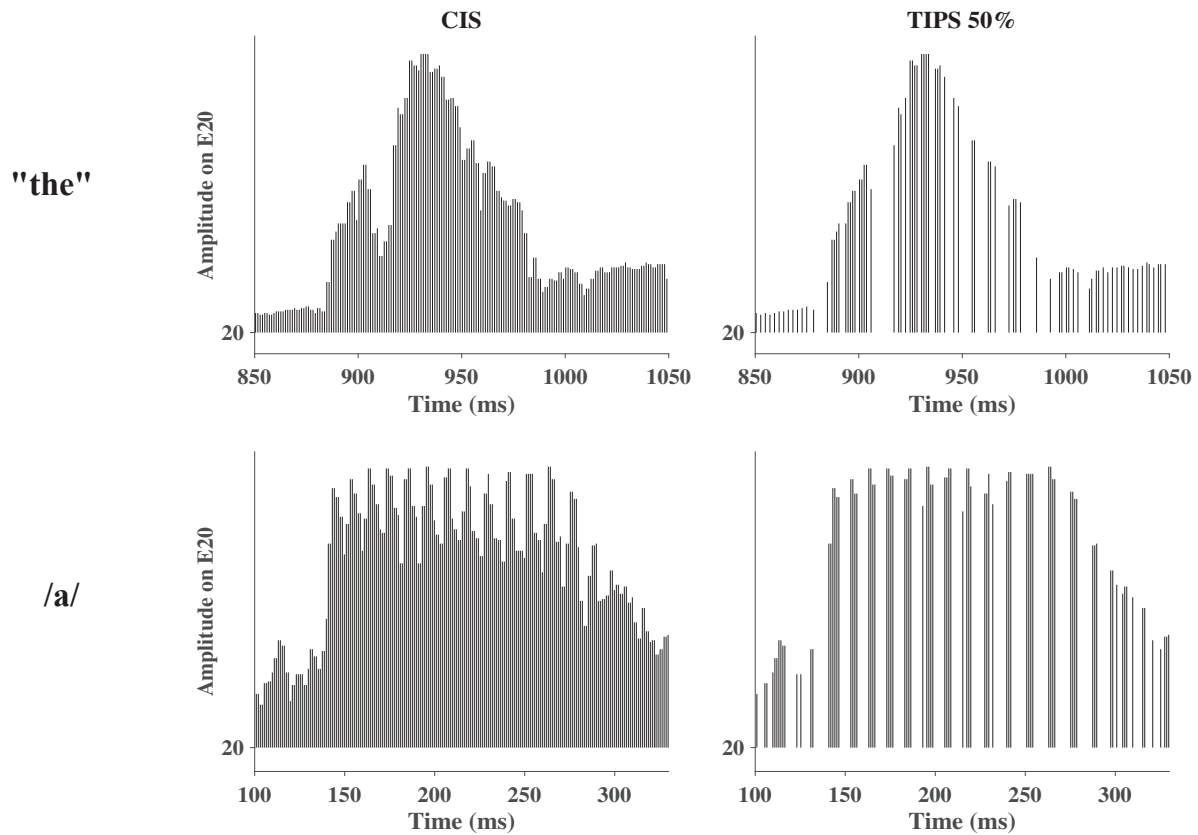


Fig. 4. Example of stimulus pattern for the syllable "the" taken from the sentence "They moved the furniture" (top) and the vowel "a" taken from the syllable "asa" with either CIS processing (left) or TIPS 50% (right). Only channel 20 is shown.

Table 1

Demographic information and details about CI devices used by the subjects. For all listeners, the inter-pulse gap was 8 μ s and the stimulation mode was MP1+2.

Subject	Sex	Age [yrs]	Implant type	Years of CI use	Strategy	Processor	Etiology	Pulse duration [μ s]	Pulse rate [pps]
S1	F	77	Freedom CI24RE Contour Advance	10	ACE	CP910	Hereditary	25	900
S2	F	66	Freedom CI24RE Contour Advance	9	ACE	CP920	Hereditary	25	1200
S3	F	49	CI532	2	ACE	CP950	Unknown	37	900
S4	F	53	CI522	1	ACE	CP950	Otosclerosis	37	900
S5	F	78	Freedom CI24RE Contour Advance	12	ACE	CP910	Otosclerosis	50	900
S6	M	76	Freedom CI24RE Contour Advance	5	ACE	CP910	Unknown	25	900
S7	M	47	Freedom CI24RE Contour Advance	13	ACE	CP950	Meningitis	37	1200
S8	M	85	Freedom CI24RE Contour Advance	6	ACE	CP1000	Hereditary	25	900

presentation, to match the number of maxima in the participant's clinical ACE settings, which was 8 out of maximally 22 for all subjects. For each subject, all other clinical settings such as pulse rate, inter-phase gap, phase duration and ground electrodes were used for the tested strategies (see Table 1), and it was ensured that electrodes selected for testing were active in the clinical map. The default CIS strategy processing settings of the NMT, e.g. those related to bandpass filtering or the envelope extraction, were used. However, the pre-emphasis filter and automatic gain control (AGC) were not incorporated into the strategy so as to minimize the influence of the pre-processing on the masking model. Sentences from the speech corpus used for testing were equalized to a fixed root mean square level of -20 dB FS. The base and saturation levels of the NMT were set to 0 and 0.39, respectively, so that the maximum output level would be at C-level while avoiding clipping of the speech stimuli. Before normalizing the levels, the speech signals and the speech-in-noise mixtures were down-sampled in Matlab to 16 kHz for compatibility with NMT processing.

2.2.3. Loudness scaling and balancing

First, the most comfortable level (MCL) was determined for each stimulus condition. To ensure that the same threshold levels (T-levels) could be used for the experimental map settings as for the regular clinical settings of the participant, the profile of T-levels from the clinical map, in clinical current units (CUs, 1 CU = 0.157 dB) was kept but set to a minimum level (lowest T-level set to zero). The global current level was then gradually increased in steps of 5 CU using 400-ms pulse trains on all 8 electrodes until the clinically set T-level was confirmed to be at an audible loudness level. As this was the case for all subjects, their clinical T-levels were kept fixed for all following experiments. For scaling the comfort levels (C-levels), the stimulus was one randomly chosen word from the speech corpus used for testing. The word was reversed in time to make it unintelligible. This was done so that the C-levels could be optimized for the stimuli from the speech corpus, without preferentially acclimatizing the subject to any specific processing strategy. Again, the same shape as for the clinical C-levels was kept, but

the difference between T- and C-level was minimized, i.e. the smallest difference between the shapes of T- and C-level was set to zero. The C-level was then gradually increased, while subjects indicated the loudness level using a chart that was marked on a scale from 0 ("off") to 10 ("too loud"). Once loudness level 7 ("loud but comfortable") was reached, the stimulus level was reduced until loudness level 6 (MCL) was confirmed.

Second, the strategies were loudness-balanced using the same time-reversed word as before. For this, again, only the C-levels were varied globally while the T-levels were kept fixed. The balancing procedure was based on that proposed by McKay and McDermott (1998) and was similar to the one used by Lamping et al. (2020) and Nogueira et al. (2016). First, the standard stimulus was presented at a fixed level followed by the signal stimulus to be adjusted. After each presentation of the stimulus pair, subjects were asked to press one of six virtual buttons on a computer screen to increase or decrease the global C-level of the signal by different amounts. This procedure was repeated until the two sounds were perceived as equally loud. The comparisons were performed twice and the average of the global C-levels was taken as the matched level. Next, standard and signal were swapped, and the previously matched level was presented as the new standard level. This procedure was repeated twice, and the average difference of the global C-levels in CUs was calculated to be the loudness balanced level. This way, the loudness levels for each of three TIPS conditions, TIPS 25%, 50% and 75%, were balanced to the CIS reference set to its MCL.

2.2.4. Speech intelligibility

Speech intelligibility in quiet and in noise was measured using the Dantale II test (Wagener et al., 2003), in which responses are entered on a virtual interface and subsequently scored automatically. The Dantale II is a closed matrix test where each sentence is composed of a name, verb, number, adjective, and noun, with 10 different options for each word, comprising a total of 50 tokens that the subject can choose from. However, it was not the case that all possible combinations of words could be presented. Rather, the corpus consists of a total of 160 recorded sentences, thereby keeping natural transitions between words. These sentences are divided into 16 lists (10 sentences per list).

Performance for speech in quiet was assessed by measuring percent-correct word scores for the loudness-balanced CIS and TIPS conditions, TIPS 25%, TIPS 50%, and TIPS 75%. Each sentence was pre-processed using the experimental map settings for the respective participant and condition. Prior to the experiment, two lists of 10 sentences each were presented in the CIS reference condition so as to provide some procedural learning. The presentation order of strategies was randomized and counterbalanced as shown in Table 2. For each processing strategy, subjects were acclimatized through presentation of sentences from the Danish HINT speech corpus (Nielsen and Dau, 2010) for about 10 min while they could read along with the printed sentence list. The HINT material was only used for acclimatization and never for testing. Directly after acclimatization, two lists of each 10 sentences from the Dantale II were presented for data collection in the respective condition and the average of the two percent-correct scores was taken as final percent correct value. A short break was provided between testing the different processing strategies. Participants were not informed of the nature of the strategies but were informed that all strategies would be different from their clinical settings, and that this might affect sound quality.

Performance for speech in noise was assessed by measuring the speech reception threshold (SRT) again for the Dantale II matrix test in an adaptive procedure that converged on 50% correct, using the speech-shaped Dantale II noise. To focus on the effect of removing pulses, and to increase the number of repetitions within one

session, testing was conducted with the CIS reference and only two of the previous TIPS conditions that had 50 and 75% of pulses removed. Each sentence was mixed with the noise using signal-to-noise ratios (SNRs) ranging from -15 to $+30$ dB, and all sentences were then pre-processed using the experimental map settings for the respective participant and condition. The same current levels as used in the speech-in-quiet experiment were kept to evaluate the different strategies. Subjects were again familiarized with each condition directly before testing by listening to speech material from the Danish HINT corpus while reading along. Thereafter, two lists of 10 Dantale II sentences, thus overall 20 sentences, were presented consecutively to measure the SRT in each condition, which was carried out twice. Presentation order of conditions was randomized and counterbalanced as far as possible and can be seen in Table 2. One adaptive run of 20 sentences was presented in noise for procedural learning prior to testing, using the CIS reference. The SNR was initially set to 0 dB and after each response, the level of the speech was held constant and the level of the noise was varied adaptively until 50% speech recognition was reached. This was achieved by adapting the SNR based on the number of words that were identified correctly (Hansen and Ludvigsen, 2001). Of the 20 sentences presented in one run, the last 15 were adjusted with ± 2 dB if all or none of the words were scored correctly, respectively, ± 1 dB when 1 or 4 words were correct and 0 dB when either 2 or 3 words were correct. For the first 5 sentences, step sizes were slightly larger ($+/- 3, 2, 1$ dB). The resulting SNR was used as the SRT and the mean of the two runs served as the final SRT score for the respective condition. A short break was offered between testing the different conditions. For both speech-in-quiet and speech-in-noise testing the participant was blinded to the experimental condition. The test was set up at the start of each condition by the experimenter, who was aware of the conditions. Nevertheless, scoring on the Dantale II test is automatic and so did not involve the experimenter.

2.2.5. Statistical analysis

For speech in quiet, statistical analysis was performed by fitting a linear mixed-effects model to the percent-correct scores. For speech in noise, a mixed model was fitted to the SRTs. For both models, stimulus condition and presentation order were set as fixed effects terms, and both models also included a subject-specific intercept, i.e. the participants were treated as a random factor. The models were implemented in R (R Core Team, 2015) using the lme4 package (Bates et al., 2015). Model selection was carried out with the lmerTest package (Kuznetsova et al., 2017), using the backward selection approach based on the stepwise deletion of model terms with high p-values (Kuznetsova et al., 2015). P-values for the fixed-effects term were calculated from F-tests (Satterthwaite's approximation of dominator degrees of freedom) while p-values for the random effects were calculated based on likelihood-ratio tests (Kuznetsova et al., 2015). The post-hoc analysis was performed through contrasts of least-square means using the emmeans library (Lenth et al., 2018; Searle et al., 1980) and the lme4 model object. The p-values were corrected for multiple comparisons using the Tukey method. Significant differences are reported using $\alpha = 0.05$.

3. Results

3.1. Loudness balancing and estimated effect on power consumption

The charge differences in the global C-level necessary to achieve the same comfortable loudness for the CIS reference and for TIPS 25, 50, and 75% can be found in Table 3. On average, charge needed

Table 2

Presentation order of conditions for the two experiments, speech in quiet and speech in noise, for each participant.

Subject	Speech in quiet				Speech in noise		
	Cond1	Cond2	Cond3	Cond4	Cond1	Cond2	Cond3
S1	CIS	TIPS 25%	TIPS 50%	TIPS 75%	CIS	TIPS 50%	TIPS 75%
S2	TIPS 75%	TIPS 50%	TIPS 25%	CIS	TIPS 50%	TIPS 75%	CIS
S3	TIPS 25%	CIS	TIPS 75%	TIPS 50%	CIS	TIPS 75%	TIPS 50%
S4	TIPS 75%	TIPS 25%	CIS	TIPS 50%	CIS	TIPS 75%	TIPS 50%
S5	TIPS 25%	TIPS 75%	TIPS 50%	CIS	TIPS 75%	CIS	TIPS 50%
S6	CIS	TIPS 50%	TIPS 75%	TIPS 25%	TIPS 50%	CIS	TIPS 75%
S7	TIPS 50%	CIS	TIPS 25%	TIPS 75%	TIPS 75%	TIPS 50%	CIS
S8	TIPS 50%	TIPS 75%	CIS	TIPS 25%	TIPS 50%	CIS	TIPS 75%

to be increased by approximately 2% (0.04 nC) for TIPS 25%, 9% (0.15 nC) for TIPS 50%, and 18% (0.29 nC) for TIPS 75%.

Power consumption in contemporary CIs can be broadly partitioned into the power used by the electronics of the sound processor and power sent to the radio frequency (RF) coil. We estimate that in modern CI systems, the power for running the sound processor is only a small part of the total power (<10%). The power sent to the RF coil consists of power needed to stimulate the electrodes, and that needed to keep the implant electronics running.

The TIPS strategy affects the overall power consumption both by reducing the number of stimulation pulses and by increasing the stimulation level per pulse. Reducing the number of pulses will reduce the power needed to stimulate the electrodes, and this power reduction will only be partly offset by the increase in the current level needed. As noted above, the stimulation current needed to be increased by 2%, 9%, and 18% when TIPS removed 25, 50, and 75% of pulses, respectively. If this component of the power current consumption drops linearly with the number of pulses and increases linearly with the stimulation level needed, then for TIPS 25%, we would expect a 25% drop (of the 90% for the RF) in power followed by a 2% increase, leading to a net power saving of 21.15%. Following the same reasoning, we would expect a net saving of 40.95 and 63.45% for TIPS 50% and TIPS 75%, respectively.

Nevertheless, the exact power savings realized by TIPS will depend on a number of aspects that will be specific to the implant type. These will include the current source circuits and the nature of the RF link. They will also be affected by the subject-specific skin flap thickness, electrode impedances and the user's listening environment. All of these factors will affect the relative contribution of the different components of the power consumption, including those that are and are not affected by the TIPS strategy. Thus, we believe that TIPS is likely to produce significant power savings but that confirmation and quantification of those savings will need to be determined using a clinical trial.

Table 3

Increase in charge of the global C-level needed to achieve the same loudness level for the TIPS conditions relative to the CIS reference for each participant, both in nano Coulomb (nC) and as a percentage (%).

Subject	TIPS 25%		TIPS 50%		TIPS 75%	
	nC	%	nC	%	nC	%
S1	0.02	1.82	0.10	9.45	0.20	17.65
S2	0.02	1.82	0.07	7.49	0.11	11.44
S3	0.00	0.00	0.32	24.20	0.62	46.12
S4	0.05	1.82	0.20	7.49	0.26	9.45
S5	0.02	1.82	0.05	5.57	0.18	19.79
S6	0.09	3.68	0.18	7.49	0.38	15.54
S7	0.09	3.68	0.14	5.57	0.33	13.48
S8	0.00	0.00	0.12	7.49	0.22	13.48
mean	0.036	1.830	0.147	9.343	0.287	18.368

3.2. Speech in quiet

Fig. 5 shows the mean percent correct scores in quiet for each individual (left) and across participants (right) where error bars depict the standard error. Performance varied markedly across subjects, ranging from ceiling effects (e.g. S1) to very low percent-correct scores (e.g. S8). The effect of condition appears to be subject-dependent, with some subjects performing on average better when fewer pulses were removed (S3, S4, S7) while others performed better with a large number of pulses removed (S5, S6, S8). On average, performance reached $75 \pm 1.74\%$ with CIS, $79 \pm 1.82\%$ with TIPS 25%, $81 \pm 1.57\%$ with TIPS 50% and $73 \pm 1.83\%$ with TIPS 75%.

There was a significant main effect of condition on percent-correct scores [$F(3,21) = 3.9$, $p = 0.024$]. Post-hoc tests with Bonferroni correction showed that performance was significantly better for TIPS 50% than for TIPS 75% ($p = 0.031$). No other comparison was statistically significant, and there was no significant effect of presentation order.

3.3. Speech in noise

Fig. 6 shows the SRT in dB SNR across CIS, TIPS 50%, and TIPS 75% for each individual (left) and for the overall mean (right). Error bars depict the standard error. Speech performance was again variable between subjects; with performance ranging from negative SRTs (e.g. S1), close to scores from normally hearing listeners (-8.4 dB SNR, Wagener et al., 2003), to highly elevated SRTs (e.g. S8). Surprisingly, S8 could achieve a 50% correct score despite results in quiet being below 50% for the CIS reference. Here, procedural learning and familiarization with the speech material across the experimental sessions might have contributed to an increase in performance for S8 in the speech-in-noise test. Examination of the adaptive tracks for S8 confirmed that, after the first reversal, the SNR decreased before converging on a stable threshold for all tested conditions.

Most subjects' SRTs improved when 50% of pulses were removed with benefits of up to 6.23 dB (S6) relative to the CIS reference, and none did worse than for the CIS reference. For some subjects, the improvement with TIPS 50% was followed by similar or worse performance relative to CIS with TIPS 75% (S3, S4, S5, S8). The mean SRT in dB SNR was 4.01 ± 1.95 for CIS, 1.59 ± 1.21 for TIPS 50%, and 4.06 ± 1.89 for TIPS 75%. There was a significant main effect of condition [$F(2,12.02) = 6.60$, $p = 0.012$] and of presentation order [$F(2,12.1) = 6.29$, $p = 0.013$]. Post-hoc comparisons corrected for multiple comparisons revealed, importantly, a significant difference between CIS and TIPS 50% ($p = 0.021$), as well as a significant difference between TIPS 50% and TIPS 75% ($p = 0.018$). For the effect of presentation order, performance was significantly worse on the third (last) than on the second condition tested ($p = 0.014$). This was potentially due to fatigue effects occurring at the end of

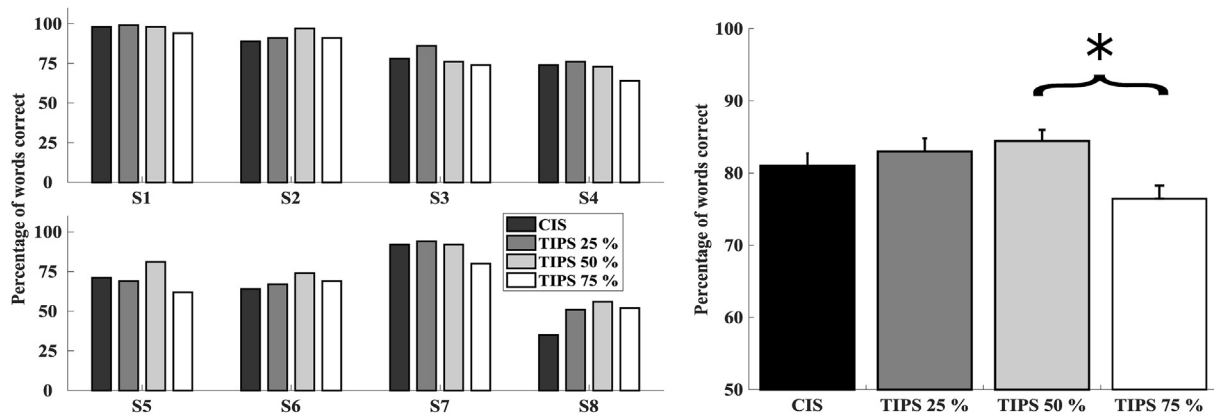


Fig. 5. Individual (left) and average (right) percentage words correct scores for CIS (black), TIPS 25% (dark grey), TIPS 50% (light grey), and TIPS 75% (white). Error bars depict the standard error.

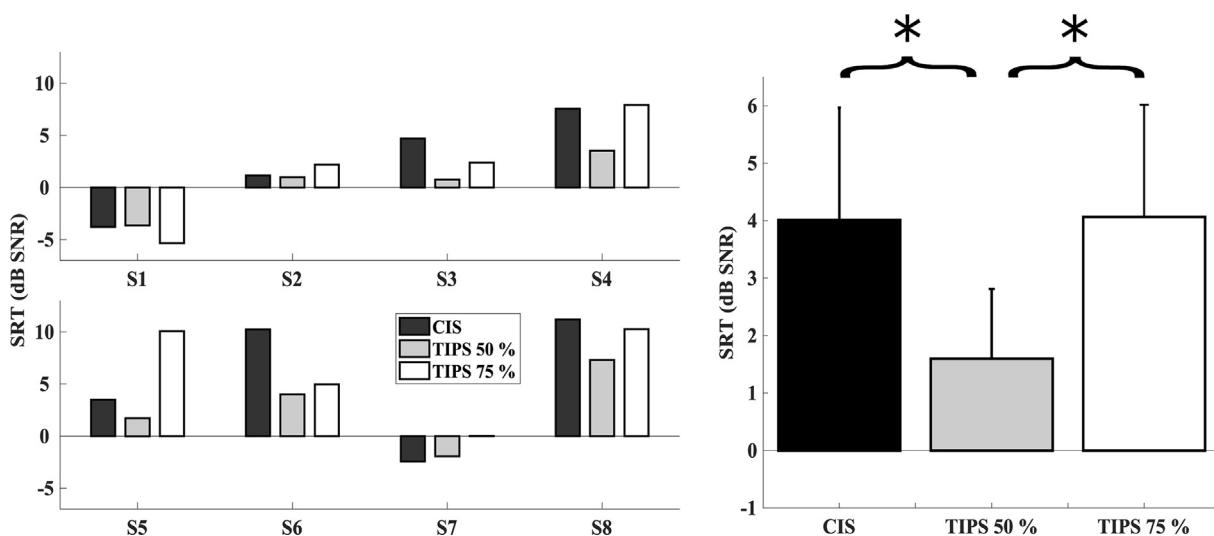


Fig. 6. Individual (left) and average (right) STR scores for CIS (black), TIPS 50% (light grey), and TIPS 75% (white). Error bars depict the standard error.

the experimental session.

4. Discussion

4.1. Subject variability and comparison to other strategies

Speech performance in quiet was similar for all TIPS conditions and the CIS reference. The TIPS condition that removed 75% of pulses did not differ significantly from CIS for speech in noise either. However, using TIPS to remove 50% of current pulses significantly improved speech intelligibility in noise, with better scores than with CIS for 6 out of 8 subjects. A post-hoc analysis of the individual results was performed to test the observation that particularly the poorer performers seemed to benefit from TIPS 50%. A significant negative correlation between the percent correct scores in quiet for CIS and the improvement in SRT over the CIS baseline when using TIPS 50% was found (Spearman's correlation coefficient $r_s = -0.71$, $p = 0.0465$) (see Fig. 7).

We can think of two possible explanations for this correlation. First, subjects with wide current spread are likely to experience more temporal interactions between pulses on different channels, thus perform poorly, and consequently benefit more when pulses are removed. Second, the poorer performers all had SRTs at positive

SNRs, as is the case for most CI users in challenging background noise conditions. Hence, the benefit from removing pulses may not solely occur from less channel interaction *per se* but also from removing pulses that are at low current levels, which are likely to be noise at positive SNRs. The 2.4 dB reduction in SRT produced when removing 50% of the pulses is also similar to that obtained by successful noise-reduction algorithms tested with CI users, which range from 2.14 dB SNR (Dawson et al., 2011) to 2.8 dB SNR (Goehring et al., 2017), at least when the masker is speech-shaped noise as used here.

A correlation analysis was performed to address whether the observed benefit in speech perception with TIPS 50% might be due to the removal of noise-dominant pulses. Fig. 8 shows the correlation coefficient values between the clean speech and the speech-in-noise mixture with decreasing SNR. The correlation was calculated for each channel, the r -values were z-transformed, averaged, and then back-transformed. This analysis was calculated for 10 sentences and the resulting correlation values were averaged to obtain the final estimate at each SNR. The black stars indicate the correlation between speech in quiet and speech in noise when processed with CIS. As expected, the correlation decreases with decreasing SNR. The red circles show correlation values using the same procedure for TIPS 50%, which seems to be more affected

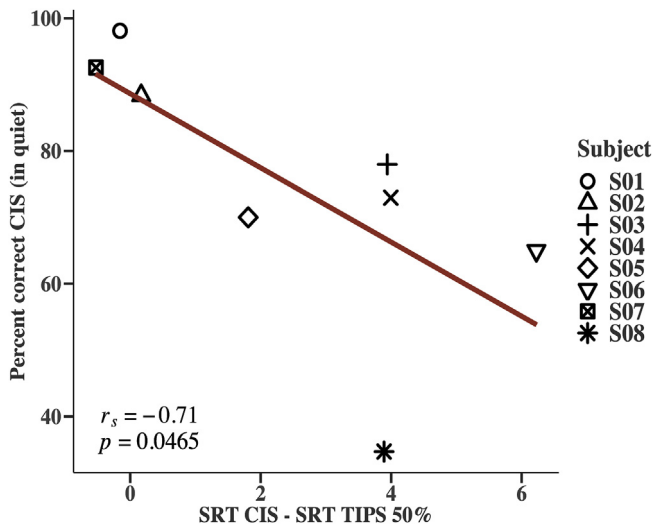


Fig. 7. Relationship between the percent correct scores for CIS in quiet and the improvement in SRT of TIPS 50% over CIS ($r_s = -0.71$, $p = 0.0465$).

when decreasing the SNR than the correlation for CIS. For comparison, the correlation value drops to 0.5 at approximately -1 and 10 dB SNR for CIS and TIPS 50%, respectively. The stronger decline in correlation values for TIPS than for CIS occurs because TIPS removes pulses from both the speech and the noise, and hence it seems unlikely that the benefit with TIPS would be caused purely by the removal of noise pulses.

In addition, and as observed in section 2.1, TIPS removed pulses located just before the onsets and just after the offsets of a sound, and/or in the troughs of modulations, as seen in Figs. 3 and 4. These “contrast enhancements” could be another reason for a benefit with TIPS in stationary background noise. For example, speech understanding by CI users can be improved by strategies that mimic adaptation, thereby enhancing slow speech modulations and the onsets and (for some strategies) offsets of sounds (Azadpour and Smith, 2016; Geurts and Wouters, 1999; Koning and Wouters, 2016). Similarly, improvements in pitch perception and, in

some instances, speech intelligibility, have been obtained with strategies that enhance modulation coding such as F0Mod (Francart et al., 2015; Laneau et al., 2006) and eTone (Vandali and van Hoesel, 2012, 2011), or that combine amplitude and rate modulation, like the ARTmod strategy (Brochier et al., 2018). For the latter, speech signals are encoded by using both amplitude and rate modulation simultaneously to enhance the temporal envelope. Interestingly, Brochier et al. (2018) also pointed out that the subjects with poorer speech understanding showed the greatest benefit from ARTmod and thus from enhanced modulation depths.

A similar approach to TIPS is the Fundamental Asynchronous Stimulus Timing (FAST) strategy (Smith et al., 2013). Here, the stimulation rate is substantially reduced by tracking the fundamental frequency and coding each amplitude modulation cycle with a single electric pulse. Preliminary results with five listeners demonstrated that speech perception was not different compared to the subject’s clinical ACE settings, but that FAST could significantly improve the detection of interaural time differences (ITDs) for bilateral CI users. As FAST reduces the number of pulses more drastically than TIPS, it might be detrimental for speech in more challenging situations. TIPS 75% did not achieve significant benefits in noise and led to mixed results with degradations in performance for some listeners. Hence, for strategies that use very sparse current patterns, the benefit of fewer pulses and less channel interactions may be confounded by the removal or distortion of speech segments that are important for intelligibility.

The improvements achieved in a previous strategy based on spectral rather than temporal masking (PACE, MP3000), were a 17% correct word score benefit with an SNR of $+15$ dB for 4 channel PACE compared to 4 channel ACE (Nogueira et al., 2005), and a 1.3 dB improvement in SRT for 4 and 8 channel PACE compared to 8 channel ACE (Buechner et al., 2008). Battery consumption could be reduced by up to 24% with a commercial version of this strategy (Buechner et al., 2011) which, however, when compared to ACE with clinical settings did not provide a benefit in speech perception. The spectral contrast enhancement algorithm (SCE, Nogueira et al., 2016), which attenuates valleys in the speech spectrum aiming to reduce spectral overlap, showed a smaller but still significant benefit of 0.57 dB when compared to ACE. Hence it appears that the benefits observed here with TIPS are roughly similar or slightly

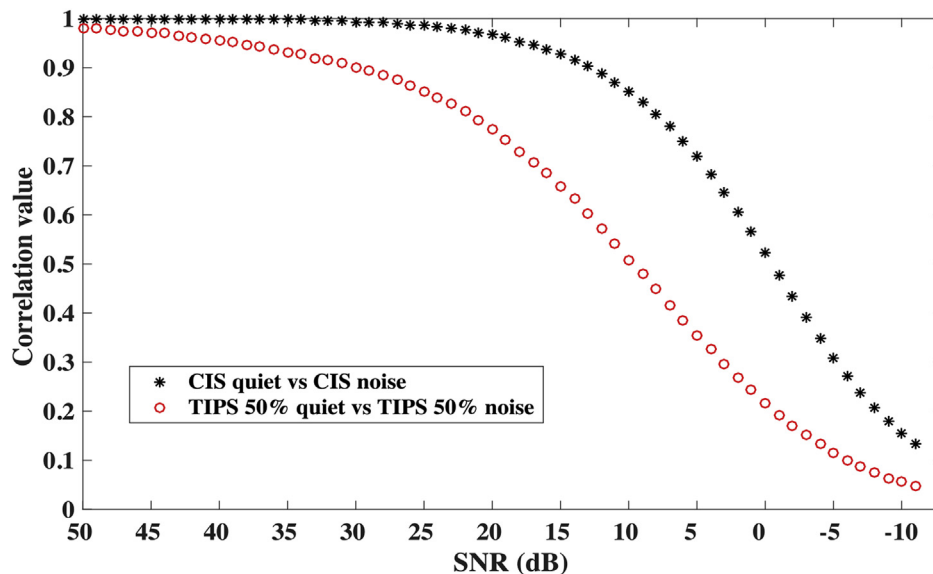


Fig. 8. Speech in quiet correlated with speech in noise over different signal-to-noise ratios for CIS (black stars) and TIPS 50% (red circles). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

larger than those observed with strategies based on spectral masking. Given the success of processing strategies based on either spectral (PACE, MP3000) or temporal (TIPS) masking, it might be worthwhile to optimize the stimulation pattern delivered to the recipient by combining them. Preliminary results from a study integrating a temporal masking processing strategy with MP3000 suggest a small benefit over MP3000 alone (Kludt et al., 2020).

Finally, across studies it seems that there is no consistent advantage of increasing or decreasing stimulation rate over the range studied here. Typically, some individuals may benefit from a specific stimulation rate, which might be connected to individual factors, such as the degree or severity of channel interaction or the state of local neural survival. However, even though no consistent effect was found across studies, individual studies found significant effects of either high- (Nie et al., 2006) or low-rate carriers (Park et al., 2012). Hence we cannot completely rule out the possibility that, for the test materials and subjects of the present study, a benefit would also have been obtained by simply reducing the pulse rate, similar to observations by Arora et al. (2009) or Brochier et al. (2017), but different to the majority of other studies mentioned above.

4.2. Perspectives

TIPS processing was applied to a CIS baseline map so that all conditions could be tested in an acute fashion. However, it would be interesting to compare TIPS to a processing strategy that is already “sparse,” such as ACE. When new processing strategies are tested acutely, one factor that needs to be considered though is the habituation of the users to their own clinical settings (Wilson and Dorman, 2008). It is unclear whether TIPS would reveal more or less benefit when applied to ACE with clinical settings (and including the usual AGC and pre-emphasis filter), as when applied to CIS. On the one hand, ACE already deselects pulses by presenting a maximum of (typically) 8 channels in each time frame. On the other hand, for the current experimental scheme only 8 electrodes dispersed across the array were active and therefore the total number of pulses was comparable to ACE. Hence, ACE and the CIS version used for this experiment may be similar in terms of the channel interaction they cause. Indeed, ACE may arguably suffer from more interactions than an 8-channel CIS strategy, as in the latter case the channels are always evenly spaced whereas with ACE adjacent channels can be stimulated concurrently. Moreover, even if TIPS does not enhance speech perception relative to ACE, only one CI manufacturer currently uses this strategy, and so TIPS would still likely provide a benefit relative to the processing strategies, more similar to CIS, that are used in other devices. However, it remains to be shown whether the benefits found with TIPS for Cochlear users here would also occur for users of CI devices from other manufacturers. Further, the comparison to or interaction with the AGC in the typical Cochlear settings are important to consider as they may be acting similarly to a strategy incorporating forward masking, albeit globally and not on a channel-by-channel basis (Vaerenberg et al., 2014).

Another aspect of interest are different types of background noise or different tasks to assess performance. For instance, contrary to steady speech-shaped noise, for modulated noise or competing speech the level of the noise in each channel will not consistently be lower or higher than the target speech at the SNRs tested. Thus, fluctuating masker types will lead to more complex spectral and temporal variations in SNR which may reduce or compromise the potential benefit of removing low-amplitude pulses. Further, the envelope contrast enhancements that occur with TIPS might, as with other strategies that enhance envelopes, lead to benefits in pitch or ITD coding (Francart et al., 2015; Laneau

et al., 2006; Smith et al., 2013). Thus, the effect of TIPS on performance in those tasks should be considered for future evaluation.

The current results reveal a range of criteria over which pulses can be removed without a substantial impact on speech performance in quiet. Nevertheless, forward masked thresholds show high variability across CI listeners (e.g. Nelson and Donaldson, 2001; Chatterjee and Kulkarni, 2017) and different criteria may yield the “best” performance for an individual. Thus, optimizing the time constant or criterion used to select current pulses on an individual level would be desirable. However, using subject-dependent criteria would be similar to finding the subject-specific carrier rate and, even if beneficial, it remains unclear whether this could be achieved in a time-efficient manner. Nevertheless, one method could be to use a spectro-temporal test, such as SMRT (Aronoff and Landsberger, 2013) or STRIPES (Archer-Boyd et al., 2018), to assess listener performance and identify the best strategy for each subject.

4.3. Benefits of TIPS

As noted above, TIPS was beneficial for speech in noise. This may be due to less channel interaction, noise removal at positive SNRs or contrast enhancements, i.e. explicit modulation coding, as well as onset and offset enhancements, or possibly a combination of all those factors.

Further, power consumption could be reduced when removing pulses, even when accounting for the charge increase needed to achieve equal loudness across strategies. CI manufacturers report that battery life with contemporary behind-the-ear processors ranges from 8 to up to 60 h, depending on whether disposable or rechargeable units are used (Wolfe and Schafer, 2015). Thus, finding a way to reduce the power consumption of CI devices is not only important for improving user convenience and to free-up resources for additional functionality, but could even lead to a reduction in the processor size which is mainly determined by the size of the battery pack.

Finally, TIPS does not require continuous tracking of the envelope, as done with e.g. FAST. Therefore, the algorithm remains computationally inexpensive and could be implemented in real-time as running temporal integrator, producing only a small processing delay. The exact computation time that would add to the delay due to the TI window (7 ms equivalent rectangular duration) will, however, depend on the processing unit.

5. Summary and conclusion

The temporal integrator, TI, model was integrated into a typical cochlear implant, CI, processing scheme and used to remove a substantial amount of pulses based on psychophysical masking. The temporal integrator processing strategy (TIPS) led to speech intelligibility in quiet that was not significantly different to the CIS baseline, even when removing up to 75% of pulses. For speech intelligibility in noise, mean SRTs in speech-shaped noise were significantly improved for TIPS when removing 50% of pulses (TIPS 50%), compared to the CIS baseline, and there was no effect on SRTs when 75% of pulses were removed (TIPS 75%). However, while TIPS 50% improved SRTs for most subjects, TIPS 75% led to more mixed results with some clear degradations in performance. Further, a significant relationship between performance scores in quiet and the improvement in SRTs with TIPS 50% indicated that poorer performers particularly benefitted from TIPS. Finally, average savings in power consumption were predicted to be at 41% with TIPS 50% for this group of users. These results indicate that TIPS may achieve improved speech perception in CI users and potential power savings that could lead to further improvements in the

usability of CI devices.

Declaration of competing interest

None.

Acknowledgments

We would like to thank our subjects for their dedicated and diligent participation. We are grateful to John Deeks and Rikke Skovhøj Sørensen for helping with the recruitment of participants and the data collection. This work was supported by the Oticon Centre of Excellence for Hearing and Speech Sciences (CHESS), by Action on Hearing Loss (Grant 82) to author TG, and by award RG91365 from the U.K. Medical Research Council to author RC.

References

- Archer-Boyd, A.W., Southwell, R.V., Deeks, J.M., Turner, R.E., Carlyon, R.P., 2018. Development and validation of a spectro-temporal processing test for cochlear-implant listeners. *J. Acoust. Soc. Am.* 144, 2983–2997. <https://doi.org/10.1121/1.5079636>.
- Aronoff, J.M., Landsberger, D.M., 2013. The development of a modified spectral ripple test. *J. Acoust. Soc. Am.* 134, EL217–EL222. <https://doi.org/10.1121/1.4813802>.
- Arora, K., Dawson, P., Dowell, R., Vandali, A., 2009. Electrical stimulation rate effects on speech perception in cochlear implants. *Int. J. Audiol.* 48, 561–567. <https://doi.org/10.1080/14992020902858967>.
- Azadpour, M., Smith, R.L., 2016. Enhancing speech envelope by integrating hair-cell adaptation into cochlear implant processing. *Hear. Res.* 342, 48–57. <https://doi.org/10.1016/j.heares.2016.09.008>.
- Balkany, T., Hodges, A., Menapace, C., Hazard, L., Driscoll, C., Gantz, B., Kelsall, D., Luxford, W., McMenomy, S., Neely, J.G., Peters, B., Pillsbury, H., Roberson, J., Schramm, D., Telian, S., Waltzman, S., Westberg, B., Payne, S., 2007. Nucleus freedom north American clinical trial. *Otolaryngol. Head Neck Surg.* 136, 757–762. <https://doi.org/10.1016/j.otohns.2007.01.006>.
- Bates, D., Mächler, M., Bolker, B.M., Walker, S.C., 2015. Fitting linear mixed-effects models using lme4. *J. Stat. Software* 67. <https://doi.org/10.18637/jss.v067.i01>.
- Bierer, J.A., Middlebrooks, J.C., 2004. Cortical responses to cochlear implant stimulation: channel interactions. *JARO J. Assoc. Res. Otolaryngol.* 5, 32–48. <https://doi.org/10.1007/s10162-003-3057-7>.
- Boulet, J., White, M.W., Bruce, I.C., 2016. Temporal considerations for stimulating spiral ganglion neurons with cochlear implants. *JARO J. Assoc. Res. Otolaryngol.* 17, 1–17. <https://doi.org/10.1007/s10162-015-0545-5>.
- Brochier, T., McDermott, H.J., McKay, C.M., 2017. The effect of presentation level and stimulation rate on speech perception and modulation detection for cochlear implant users. *J. Acoust. Soc. Am.* 141, 4097–4105. <https://doi.org/10.1121/1.4983658>.
- Brochier, T., McKay, C., McDermott, H., 2018. Encoding speech in cochlear implants using simultaneous amplitude and rate modulation. *J. Acoust. Soc. Am.* 144, 2042–2051. <https://doi.org/10.1121/1.5055989>.
- Buechner, A., Beynon, A., Szyfter, W., Niemczyk, K., Hoppe, U., Hey, M., Brokx, J., Eyles, J., Van de Heyning, P., Paludetti, G., Zarowski, A., Quaranta, N., Wesarg, T., Festen, J., Olze, H., Dhooge, I., Müller-Deile, J., Ramos, A., Roman, S., Piron, J.-P., Cuda, D., Burdo, S., Grolman, W., Vailland, S.R., Huarte, A., Frachet, B., Morera, C., García-Ibáñez, L., Abels, D., Walger, M., Müller-Mazotta, J., Leone, C.A., Meyer, B., Dillier, N., Steffens, T., Gentine, A., Mazzoli, M., Rypkema, G., Killian, M., Smoorenburg, G., 2011. Clinical evaluation of cochlear implant sound coding taking into account conjectural masking functions, MP3000™. *Cochlear Implants Int.* 12, 194–204. <https://doi.org/10.1179/1754762811Y00000000009>.
- Buechner, A., Nogueira, W., Edler, B., Battmer, R.D., Lenarz, T., 2008. Results from a psychoacoustic model-based strategy for the nucleus-24 and freedom cochlear implants. *Otol. Neurotol.* 29, 189–192. <https://doi.org/10.1097/mao.0b013e318162512c>.
- Carlyon, R.P., van Wieringen, A., Deeks, J.M., Long, C.J., Lyzenga, J., Wouters, J., 2005. Effect of inter-phase gap on the sensitivity of cochlear implant users to electrical stimulation. *Hear. Res.* 205, 210–224. <https://doi.org/10.1016/j.heares.2005.03.021>.
- Chatterjee, M., Kulkarni, A.M., 2017. Recovery from forward masking in cochlear implant listeners depends on stimulation mode, level, and electrode location. *J. Acoust. Soc. Am.* 141, 3190–3202. <https://doi.org/10.1121/1.4983156>.
- Cosentino, S., Deeks, J.M., Carlyon, R.P., 2015. Procedural factors that affect psychophysical measures of spatial selectivity in cochlear implant users. *Trends Hear* 19, 1–16. <https://doi.org/10.1177/2331216515607067>.
- Dawson, P.W., Mauger, S.J., Hersbach, A. a., 2011. Clinical evaluation of signal-to-noise ratio-based noise reduction in Nucleus® cochlear implant recipients. *Ear Hear.* 32, 382–390. <https://doi.org/10.1097/AUD.0b013e318201c200>.
- Dorman, M.F., Loizou, P.C., Rainey, D., 1997. Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. *J. Acoust. Soc. Am.* 102, 2403–2411. <https://doi.org/10.1121/1.419603>.
- Dynes, S.B.C., Delgutte, B., 1992. Phase-locking of auditory-nerve discharges to sinusoidal electric stimulation of the cochlea. *Hear. Res.* 58, 79–90. [https://doi.org/10.1016/0378-5955\(92\)90011-B](https://doi.org/10.1016/0378-5955(92)90011-B).
- Francart, T., Osses, A., Wouters, J., 2015. Speech perception with F0mod, a cochlear implant pitch coding strategy. *Int. J. Audiol.* 54, 424–432. <https://doi.org/10.3109/14992027.2014.989455>.
- Friesen, L.M., Shannon, R.V., Baskent, D., Wang, X., 2001. Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. *J. Acoust. Soc. Am.* 110, 1150–1163. <https://doi.org/10.1121/1.1381538>.
- Friesen, L.M., Shannon, R.V., Cruz, R.J., 2005. Effects of stimulation rate on speech recognition with cochlear implants. *Audiol. Neurotol.* 10, 169–184. <https://doi.org/10.1159/000084027>.
- Fu, Q.J., Nogaki, G., 2005. Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing. *JARO J. Assoc. Res. Otolaryngol.* 6, 19–27. <https://doi.org/10.1007/s10162-004-5024-3>.
- Geurts, L., Wouters, J., 1999. Enhancing the speech envelope of continuous interleaved sampling processors for cochlear implants. *J. Acoust. Soc. Am.* 105, 2476–2484. <https://doi.org/10.1121/1.426851>.
- Goehring, T., Bolner, F., Monaghan, J.J.M., van Dijk, B., Zarowski, A., Bleeck, S., 2017. Speech enhancement based on neural networks improves speech intelligibility in noise for cochlear implant users. *Hear. Res.* 344, 183–194. <https://doi.org/10.1016/j.heares.2016.11.012>.
- Hansen, M., Ludvigsen, C., 2001. Dantale II. Danske Hagerman Sætninger. Danske tale audiometrimaterialer.
- Heffer, L.F., Sly, D.J., Fallon, J.B., White, M.W., Shepherd, R.K., O'Leary, S.J., 2010. Examining the auditory nerve fiber response to high rate cochlear implant stimulation: chronic sensorineural hearing loss and facilitation. *J. Neurophysiol.* 104, 3124–3135. <https://doi.org/10.1152/jn.00500.2010>.
- Kiefer, J., von Ilberg, C., Rupprecht, V., Hubner-Egner, J., Knecht, R., 2000. Optimized speech understanding with the continuous interleaved sampling speech coding strategy in patients with cochlear implants: effect of variations in stimulation rate and number of channels. *Ann. Otol. Rhinol. Laryngol.* 109, 1009–1020. <https://doi.org/10.1177/000348940010901105>.
- Kludt, E., Nogueira, W., Lenarz, T., Buechner, A., 2020. Integration of temporal masking into the MP3000 coding strategy. *PsyArXiv*. <https://psyarxiv.com/gvntj/>.
- Koning, R., Wouters, J., 2016. Speech onset enhancement improves intelligibility in adverse listening conditions for cochlear implant users. *Hear. Res.* 342, 13–22. <https://doi.org/10.1016/j.heares.2016.09.002>.
- Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., 2017. lmerTest package: tests in linear mixed effects models. *J. Stat. Software* 82. <https://doi.org/10.18637/jss.v082.i13>.
- Kuznetsova, A., Christensen, R.H.B., Bavay, C., Brockhoff, P.B., 2015. Automated mixed ANOVA modeling of sensory and consumer data. *Food Qual. Prefer.* 40, 31–38. <https://doi.org/10.1016/j.foodqual.2014.08.004>.
- Lamping, W., Deeks, J.M., Marozeau, J., Carlyon, R.P., 2020. The effect of phantom stimulation and pseudomonophasic pulse shapes on pitch perception by cochlear implant listeners. *PsyArXiv*. <https://psyarxiv.com/ugkem>.
- Laneau, J., Moonen, M., Wouters, J., 2006. Factors affecting the use of noise-band vocoders as acoustic models for pitch perception in cochlear implants. *J. Acoust. Soc. Am.* 119, 491–506. <https://doi.org/10.1121/1.2133391>.
- Langner, F., Saoji, A.A., Büchner, A., Nogueira, W., 2017. Adding simultaneous stimulating channels to reduce power consumption in cochlear implants. *Hear. Res.* 345, 96–107. <https://doi.org/10.1016/j.heares.2017.01.010>.
- Lenth, R., Singmann, H., Love, J., Buurkner, P., Herve, M., 2018. Estimated Marginal Means, aka Least-Squares Means. License. <https://CRAN.R-project.org/package=emmeans>.
- Loizou, P.C., Poroy, O., Dorman, M., 2000. The effect of parametric variations of cochlear implant processors on speech understanding. *J. Acoust. Soc. Am.* 108, 790–802. <https://doi.org/10.1121/1.429612>.
- Marozeau, J., McDermott, H.J., Swanson, B.A., McKay, C.M., 2015. Perceptual interactions between electrodes using focused and monopolar cochlear stimulation. *JARO J. Assoc. Res. Otolaryngol.* 16, 401–412. <https://doi.org/10.1007/s10162-015-0511-2>.
- McDermott, H.J., McKay, C.M., Vandali, A.E., 1992. A new portable sound processor for the University of Melbourne/Nucleus Limited multielectrode cochlear implant. *J. Acoust. Soc. Am.* 91, 3367–3371. <https://doi.org/10.1121/1.402826>.
- McKay, C.M., Henshall, K.R., 2010. Amplitude modulation and loudness in cochlear implantees, 111, pp. 101–111. <https://doi.org/10.1007/s10162-009-0188-5>.
- McKay, C.M., Henshall, K.R., Hull, A.E., 2005. The effect of rate of stimulation on perception of spectral shape by cochlear implantees. *J. Acoust. Soc. Am.* 118, 386–392. <https://doi.org/10.1121/1.1937349>.
- McKay, C.M., Lim, H.H., Lenarz, T., 2013. Temporal processing in the auditory system: insights from cochlear and auditory midbrain implantees. *JARO J. Assoc. Res. Otolaryngol.* 14, 103–124. <https://doi.org/10.1007/s10162-012-0354-z>.
- McKay, C.M., McDermott, H.J., 1998. Loudness perception with pulsatile electrical stimulation: the effect of interpulse intervals. *J. Acoust. Soc. Am.* 104, 1061–1074. <https://doi.org/10.1121/1.423316>.
- McKay, C.M., McDermott, H.J., 1996. The perception of temporal patterns for electrical stimulation presented at one or two intracochlear sites, 100, 1081–1092.
- McKay, C.M., Remine, M.D., McDermott, H.J., 2001. Loudness summation for pulsatile electrical stimulation of the cochlea: effects of rate, electrode separation,

- level, and mode of stimulation. *J. Acoust. Soc. Am.* 110, 1514–1524. <https://doi.org/10.1121/1.1394222>.
- Middlebrooks, J.C., 2004. Effects of cochlear-implant pulse rate and inter-channel timing on channel interactions and thresholds. *J. Acoust. Soc. Am.* 116, 452–468. <https://doi.org/10.1121/1.1760795>.
- Monaghan, J.J.M., Carlyon, R.P., Deeks, J.M., 2019. Amplitude modulation depth discrimination by cochlear implant users. *Conf. Implant. Audit. Prostheses*.
- Moore, B.C.J., 2007. Temporal resolution and temporal integration. In: *Cochlear Hearing Loss - Physiological, Psychological and Technical Issues*. John Wiley & Sons Ltd, pp. 117–142.
- Moore, B.C.J., Peters, R.W., Glasberg, B.R., 1996. Detection of decrements and increments in sinusoids at high overall levels. *J. Acoust. Soc. Am.* 97 <https://doi.org/10.1121/1.413019>, 3329–3329.
- Nelson, D.A., Donaldson, G.S., 2001. Psychophysical recovery from single-pulse forward masking in electric hearing. *J. Acoust. Soc. Am.* 109, 2921–2933. <https://doi.org/10.1121/1.1514935>.
- Nie, K., Barco, A., Zeng, F.G., 2006. Spectral and temporal cues in cochlear implant speech perception. *Ear Hear.* 27, 208–217. <https://doi.org/10.1097/01.aud.0000202312.31837.25>.
- Nielsen, J.B., Dau, T., 2010. The Danish hearing in noise test. *Int. J. Audiol.* 50, 202–208. <https://doi.org/10.3109/149920207.2010.524254>.
- Nogueira, W., Buechner, A., Lenarz, T., Edler, B., 2005. A Psychoacoustic “NofM”-Type Speech Coding Strategy for Cochlear Implants, pp. 3044–3059.
- Nogueira, W., Rode, T., Buechner, A., 2016. Spectral contrast enhancement improves speech intelligibility in noise for cochlear implants. *J. Acoust. Soc. Am.* 728–739. <https://doi.org/10.1121/1.4939896>.
- Oxenham, A.J., 2001. Forward masking: adaptation or integration? *J. Acoust. Soc. Am.* 109, 732–741. <https://doi.org/10.1121/1.1336501>.
- Oxenham, A.J., Moore, B.C.J., 1994. Modeling the additivity of nonsimultaneous masking. *Hear. Res.* 80, 105–118. [https://doi.org/10.1016/0378-5955\(94\)90014-0](https://doi.org/10.1016/0378-5955(94)90014-0).
- Park, S.H., Kim, E., Lee, H.J., Kim, H.J., 2012. Effects of electrical stimulation rate on speech recognition in cochlear implant users. *Korean J. Audiol.* 16, 6–9. <https://doi.org/10.7874/kja.2012.16.1.6>.
- Plack, C.J., Moore, B.C.J., 1990. Temporal window shape as a function of frequency and level. *J. Acoust. Soc. Am.* 87, 2178–2187. <https://doi.org/10.1121/1.399185>.
- Plack, C.J., Oxenham, A.J., 2002. Linear and nonlinear processes in temporal masking. *Hear. Res.* 88, 348–358.
- Plack, C.J., Oxenham, A.J., 1998. Basilar-membrane nonlinearity and the growth of forward masking. *J. Acoust. Soc. Am.* 103, 1598–1608. <https://doi.org/10.1121/1.421294>.
- Plack, C.J., Oxenham, A.J., Drga, V., 2002. Linear and nonlinear processes in temporal masking. *Acta Acust. united with Acust.* 88 (3), 348–358.
- Plant, K., Holden, L., Skinner, M., Arcaroli, J., Whitford, L., Law, M.-A., Nel, E., 2007. Clinical evaluation of higher stimulation rates in the nucleus research platform 8 system. *Ear Hear.* 28, 381–393. <https://doi.org/10.1097/AUD.0b013e31804793ac>.
- Plant, K.L., Whitford, L.A., Psarros, C.E., Vandali, A.E., 2002. Parameter selection and programming recommendations for the ACE and CIS speech-processing strategies in the Nucleus 24 cochlear implant system. *Cochlear Implants Int.* 3, 104–125. <https://doi.org/10.1179/cim.2002.3.2.104>.
- R Core Team, 2015. A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. <https://doi.org/10.1093/ije/duy041>. <https://www.r-project.org/>.
- Rubinstein, J.T.Y., Wilson, B.S., Finley, C.C., Abbas, P.J., 1999. Pseudospontaneous activity: stochastic independence of auditory nerve fibers with electrical stimulation. *Hear. Res.* 127, 108–118.
- Searle, S.R., Speed, F.M., Milliken, G.A., 1980. Population marginal means in the linear model: an alternative to least squares means. *Am. Statistician* 34, 216–221. <https://doi.org/10.1080/00031305.1980.10483031>.
- Shannon, R.V., Fu, Q., Galvin, J., Friesen, L., 2004. Speech perception with cochlear implants. In: Zeng, F., Popper, A.N., Richard, R.F. (Eds.), *Cochlear Implants: Auditory Prostheses and Electric Hearing*. Springer Handbook of Auditory Research. Springer, New York, NY.
- Shannon, R.V., 1989. A model of threshold for pulsatile electrical stimulation of cochlear implants. *Hear. Res.* 40, 197–204. [https://doi.org/10.1016/0378-5955\(89\)90160-3](https://doi.org/10.1016/0378-5955(89)90160-3).
- Shannon, R.V., Cruz, R.J., Galvin, J.J., 2011. Effect of stimulation rate on cochlear implant users' phoneme, word and sentence recognition in quiet and in noise. *Audiol. Neurotol.* 16, 113–123. <https://doi.org/10.1159/000315115>.
- Shannon, R.V., 1983. Multichannel electrical stimulation of the auditory nerve in man. II. Channel interaction. *Hear. Res.* 12, 1–16.
- Skinner, M.W., Holden, L.K., Whitford, L.A., Plant, K.L., Psarros, C., Holden, T.A., 2002. Speech recognition with the Nucleus 24 SPEAK, ACE, and CIS speech coding strategies in newly implanted adults. *Ear Hear.* 23, 207–223. <https://doi.org/10.1097/00003446-200206000-00005>.
- Smith, Z.M., Parkinson, W.S., Krishnamoorthi, H., 2013. Efficient coding for auditory prostheses. *Conf. Implant. Audit. Prostheses*.
- Swanson, B., Mauch, H., 2006. Nucleus MATLAB Toolbox, vol. 4, p. 20.
- Townshend, B., Cotter, N., Van Compernelle, D., White, R.L., 1987. Pitch perception by cochlear implant subjects. *J. Acoust. Soc. Am.* 82, 106–115. <https://doi.org/10.1121/1.395554>.
- Vaerenberg, B., Govaerts, P., Stainsby, T., Nopp, P., Gault, A., Gnansia, D., 2014. A uniform graphical representation of intensity coding in current generation cochlear implant systems. *Ear Hear.* 35, 533–543. <https://doi.org/10.1097/AUD.0000000000000039>.
- Vandali, A.E., van Hoesel, R.J.M., 2012. Enhancement of temporal cues to pitch in cochlear implants: effects on pitch ranking. *J. Acoust. Soc. Am.* 132, 392–402. <https://doi.org/10.1121/1.4718452>.
- Vandali, A.E., van Hoesel, R.J.M., 2011. Development of a temporal fundamental frequency coding strategy for cochlear implants. *J. Acoust. Soc. Am.* 129, 4023–4036. <https://doi.org/10.1121/1.3573988>.
- Vandali, A.E., Whitford, L.A., Plant, K.L., Clark, G.M., 2000. Speech perception as a function of electrical stimulation rate: using the Nucleus 24 cochlear implant system. *Ear Hear.* 21, 608–624.
- Verschuur, C.A., 2005. Effect of stimulation rate on speech perception in adult users of the Med-El CIS speech processing strategy. *Int. J. Audiol.* 44, 58–63. <https://doi.org/10.1080/14992020400022488>.
- Wagener, K., Josvassen, J.L., Ardenkjær, R., 2003. Design, optimization and evaluation of a Danish sentence test in noise. *Int. J. Audiol.* 42, 10–17. <https://doi.org/10.3109/14992020309056080>.
- Weber, B.P., Lai, W.K., Dillier, N., Von Wallenberg, E.L., Killian, M.J.P., Pesch, J., Battmer, R.D., Lenarz, T., 2007. Performance and preference for ACE stimulation rates obtained with nucleus RP 8 and freedom system. *Ear Hear.* 28, 46–48. <https://doi.org/10.1097/AUD.0b013e3180315442>.
- Wilson, B.S., Dorman, M.F., 2008. Cochlear implants: a remarkable past and a brilliant future. *Hear. Res.* 242 (1–2), 3–21. <https://doi.org/10.1016/j.heares.2008.06.005>.
- Wilson, B.S., Finley, C.C., Lawson, D.T., 1988. Comparative studies of speech processing strategies for cochlear implants. *Laryngoscope* 98.
- Wilson, B.S., Finley, C.C., Lawson, D.T., Wolford, R.D., Eddington, D.K., Rabinowitz, W.M., 1991. Better speech recognition with cochlear implants. *Nature* 352, 236–238. <https://doi.org/10.1038/352236a0>.
- Wolfe, J., Schafer, E.C., 2015. *Programming Cochlear Implants*, 2nd revise. Plural Publishing Inc, San Diego, CA.
- Wouters, J., McDermott, H.J., Francart, T., 2015. Sound coding in cochlear implants. *IEEE Signal Process. Mag.* 32, 67–80. <https://doi.org/10.1109/MSP.2014.2371671>.
- Zeng, F.G., Rebscher, S., Harrison, W., Sun, X., Feng, H., 2008. Cochlear implants: system design, integration, and evaluation. *IEEE Rev. Biomed. Eng.* 1, 115–142. <https://doi.org/10.1109/RBME.2008.2008250>.